因子分析の数値実験によるApproach (統計的多変量解析の研究報告集)

福富, 和夫

数理解析研究所講究録
44巻 1968年 69-78

因子分析の数値実験による Approach

芝浦工大　福富和夫

§1　序

数値実験による因子分析の研究はすでに Wold [7], Jöreskog [4], Cliff & Pennell [2] 等により因子抽出法の比較，サンプルサイズ，loading の大きさ，Communality の大きさ等と loading の推定量の分散の関係，分布の歪みが推定値に与える影響などについて種々行われてきているが，ここでは因子数について検討するため一連の数値実験を試みた。

まず因子分析のモデルとして次のものを考える。

(1)　$\underset{\sim}{x} = \underset{\sim}{\mu} + F\underset{\sim}{f} + V\underset{\sim}{e}$

$\underset{\sim}{x}$ は p 次の縦ベクトル，F は因子行列で $p \times m$，$V\underset{\sim}{e}$ は Specific 及び Error の項で V は対角行列

さらにモデルに一意性を与える為　$E(\underset{\sim}{f}) = \underset{\sim}{0}$，$E(\underset{\sim}{e}) = \underset{\sim}{0}$

$E(\underset{\sim}{f}\underset{\sim}{f}')=I_m$, $E(\underset{\sim}{e}\underset{\sim}{e}')=I_p$, $E(\underset{\sim}{f}\underset{\sim}{e}')=0$ を仮定する。

ここでサンプルの分散共分散行列（または相関行列）から共通因子数mを既知としてFおよびVの推定値を求めるのであるが，問題はmの値である。

Cattell〔1〕，Guttman〔3〕が主張するように実態科学特に心理学の立場からすれば非常に多くの因子が存在するものとみる妥当であろう。（Cattellは因子をtrue factorとerror factorに分けて後者はサンプリングにより作り出されたものとしている。更にtrue factorの中でも多くのtrivialなものを無視して影響の大きい因子だけを取扱うとしている。）

一方有意性の検定などによりmを決めるとテスト数Pに比してかなり小さい値が得られるのが普通である。さらにmに大きな値（$P-1/2<m$）を与えると解が一意でなくなる場合がある（Thurstone〔6〕）といわれているが，われわれの実験でも $\frac{P-1}{2}<m$ の場合は全て初期値の与え方によりiterationがいくつかの異なる値に収束するという結果が得られている。（ここで解の一意性と収束の一意性とは別であることに注意。収束については何の結果も得られていない。（Rao〔5〕参照）実際 $\frac{P-1}{2}\geqq m$ の場合でも初期値により異なる値に収束した例もある。（表3））

したがって因子数として真の値より小さい値を与えること

になる場合が多い。

いま上のモデルにおいて　T をある m 次の直交行列とすると (1) 式は

$$\underset{\sim}{x} = \underset{\sim}{\mu} + F T T' \underset{\sim}{f} + V \underset{\sim}{e}$$

T を (T_1, T_2) (T_1 は $m \times k$, T_2 は $m \times (m-k)$ の行列) に分割して

$$T'\underset{\sim}{f} = \begin{pmatrix} T_1'\underset{\sim}{f} \\ T_2'\underset{\sim}{f} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{f}_1^* \\ \underset{\sim}{f}_2^* \end{pmatrix}$$ とおくと

(2) $$\underset{\sim}{x} = \underset{\sim}{\mu} + F_1^* \underset{\sim}{f}_1^* + F_2^* \underset{\sim}{f}_2^* + V \underset{\sim}{e}$$

ここで　$F_1^* = F T_1$, $F_2^* = F T_2$。

よって真の因子数 m より小さい値 k を因子数と指定した時因子抽出の Criterion は

$$\min_{T_1} \| \mathrm{nondiag}\,(R - F T_1 T_1' F') \|^2$$ であるべきであろう。

ただし　$\| \mathrm{nondiag}\, A \|^2 = \sum_{i \neq j} a_{ij}^2$,　T_1 は縦ベクトルが orthonormal な $m \times k$ の行列。

この場合得られる推定値のCommunalityは明らかに真のCommunalityより大きくはならない。

ところで実際に行われる因子抽出のCriterionは，P.F.A.では

$$\min_{\tilde{F}} \|\mathrm{nondiag}\,(R-\tilde{F}\tilde{F}')\|^2$$

またC.F.A.では

$$\min_{\tilde{F},\tilde{V}} \|\mathrm{nondiag}\,\tilde{V}^{-2}(R-\tilde{F}\tilde{F}')\,\tilde{V}^{-2}\|^2 \quad \text{である。}$$

（$\tilde{F}$は $p\times k$ の行列，$\tilde{V}$は要素が正の対角行列）。

実験の結果は真の因子数を指定した時はこれらの方法（ここではC.F.A.を用いた。）で正しい因子行列とみられるものが得られたが，真の因子数より小さい値を指定した時は無縁と思われる因子がでてきており、得られたCommunalityも真の値より大きくなるという結果もでている。

次にfactorial invarianceについて考えてみると、あるテストの組$\underset{\sim}{x}_1$に新しいテストの組$\underset{\sim}{x}_2$を加えた場合共通因子の数が変らなければモデルは

$$\begin{pmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{pmatrix} + \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} \underset{\sim}{f} + \begin{pmatrix} V_1 & \\ & V_2 \end{pmatrix} \begin{pmatrix} \underset{\sim}{e}_1 \\ \underset{\sim}{e}_2 \end{pmatrix}$$

となりfactorial invarianceは成立つであろうが、因子数が増

加する場合即ち z_2 に別の新たな因子が含まれているが、前にはSpecificであったものが共通因子に変る場合は前と同じ因子数を指定したのではfactorial invarianceは成立たないであろう。

§2. 実験の方法

まず真の因子行列 F を与える。この時 FF' の対角要素が1より小さくなるようにして $FF' - \mathrm{diag}\,FF' + I$ を相関行列 R としこれにLawleyの方法によるC.F.A.を用いて因子抽出を行った。この際必要とする初期値は普通はセントロイド法により求め、特にこれと異なる初期値を要するときは主成分を用いた。Iterationの収束はCommunalityの変化が全て .002 以下になったとき打切っている。Varimax回転のcycle数は m 次の行列では $m(m-1)/2$ 回で止るようになっている。

またの真の因子行列 F と推定値の行列 $\hat{F}$ との比較が意味を持つ即ち同じ型の行列のときは

$$\min_{T} \| \hat{F} T - F \|^2 \qquad (T \text{ は直交行列})$$

を用いてfitさせて比較した。（この回転はVarimax回転と同様の方法で得られ、Iterationのcycleはcriterionの変化が .001 以下になったとき打切った。）

以上の計算に使用したコンピューターはIBM 1620である。

§3. 実験とその結果

表1Aの因子行列は1〜6のテストに対しては初めの二列がその因子行列で他はspecific，1〜8のテストに対しては初めの三列，1〜10のテストに対しては初めの四列が因子行列になるように作られている。これらの因子行列から作られた相関行列に対し正しい因子数を指定して因子抽出を行った結果が表2Aである。この場合は真の因子行列が再現されたとみてよい。これに対し表1Aの因子行列にいくつかの要素を加えて作った因子行列（表1B）の場合は（下線のあるものが新たに加えられた要素）1〜6のテストでは4列と6列のみがspecificで因子数は6となり，1〜8及び1〜10のテストでは全ての列が共通因子となって因子数8となっている。これらにそれぞれ2, 3, 4の因子数を指定して因子抽出を行った結果が表2Bである。この場合は特に$P=10$でテスト1及び9のcommunalityが真のものより著しく大きくなっている点に注意したい。

表3は8×4の因子行列からの相関行列について初期値をセントロイドと主成分で与えたときのC.F.A.による解のcommunalityを比較したものである。こゝでは$P=8$だから$m=4$では$P-1/2<m$が成立つ場合である。これ以外でも$P-1/2<m$の場合

表 1A

TRUE FACTOR MATRIX A

	1	2	3	4	5	6	7	8	COMM. 2	COMM. 3	COMM. 4
1	.65							.46	.4225	.4225	.4225
2	.62	.49							.6245	.6245	.6245
3	.59								.3481	.3481	.3481
4		.43	.64			.48			.1849	.5945	.5945
5	.63						.47		.3969	.3969	.3969
6		.40			.62				.1600	.1600	.1600
7			.60	.48						.3600	.5904
8		.42	.61							.5485	.5485
9			.58	.49							.5765
10		.50	.54								.5416

表 1B

TRUE FACTOR MATRIX B

	1	2	3	4	5	6	7	8	COMM. 2	COMM. 3	COMM. 4
1	.65						.37	.46	.7710	.7710	.7710
2	.62	.49	.35						.7470	.7470	.7470
3	.59			.41		.48			.3481	.7466	.7466
4		.43	.64					.32	.6969	.6969	.6969
5	.63				.38		.47		.7622	.7622	.7622
6		.40			.62				.5444	.5444	.5444
7			.60	.48						.5904	.5904
8		.42	.61							.5485	.5485
9			.58	.49		.42					.7529
10		.50	.54								.5416

表 3

		m = 4		m = 2	
	TRUE COMM.	CENT.	PRIN.	CENT.	PRIN.
1	.5776	.5784	.5774	.0012	.5759
2	.6084	.6073	.6077	.0012	.6048
3	.5737	.5755	.4821	.1262	.4319
4	.3969	.3976	.8855	.2095	.1935
5	.7818	.7753	.7696	.7867	.7990
6	.5953	.5993	.5993	.5955	.5886
7	.3481	.4895	.6797	.3357	.0003
8	.7489	.5639	.4029	.6962	.0514

表2A

	COMM.	F1	F2
1	.4232	.6505	.0022
2	.6109	.6106	.4802
3	.3484	.5903	.0022
4	.1901	-.0003	.4360
5	.3974	.6304	.0022
6	.1635	-.0000	.4043

	COMM.	F1	F2	F3
1	.4227	.6501	.0008	-.0005
2	.6191	.6187	.4860	.0018
3	.3482	.5901	.0008	-.0005
4	.5942	.0000	.4306	.6393
5	.3970	.6301	.0008	-.0005
6	.1615	-.0000	.4019	-.0011
7	.3620	.0001	-.0016	.6017
8	.5486	.0000	.4208	.6095

	COMM.	F1	F2	F3	F4
1	.4227	.6501	.0057	-.0049	.0060
2	.6014	.6155	.4713	.0134	-.0162
3	.3482	.5900	.0052	-.0045	.0055
4	.5942	.0006	.4210	.6404	-.0003
5	.3971	.6301	.0056	-.0048	.0059
6	.1776	-.0036	.4208	-.0151	.0183
7	.5904	-.0000	-.0000	.5999	.4801
8	.5484	.0006	.4191	.6105	-.0004
9	.5764	-.0000	.0001	.5797	.4902
10	.5402	.0007	.5007	.5399	.0000

注：抽出後 直交回転により真の因子行列にfitしたもの

表2B

	COMM.	F1	F2
1	.5171	.6872	-.2118
2	.5168	.4521	-.5589
3	.2683	.5119	.0793
4	.7052	-.0725	.8366
5	.6460	.8003	.0744
6	.0619	.1395	.2060

	COMM.	F1	F2	F3
1	.5885	.7650	.0505	.0241
2	.5524	.5140	.5073	.1755
3	.2929	.5349	.0540	.0670
4	.5935	.0699	.7600	-.1050
5	.6884	.7618	-.0872	-.3170
6	.4672	.0362	.1034	.6747
7	.2431	.0478	.4864	.0652
8	.5779	-.0098	.7414	-.1673

	COMM.	F1	F2	F3	F4
1	.8873	.9232	-.0497	-.1160	-.1379
2	.6972	.4728	.1403	-.4118	.5332
3	.6303	.5005	.5435	.1927	.2173
4	.6615	.0786	.1682	.7877	.0806
5	.4830	.6635	.0221	.0529	.2018
6	.1200	.0364	-.0461	-.1632	.2999
7	.4389	-.0196	.5420	.3796	-.0254
8	.5485	-.0377	.1976	.6710	.2397
9	.8385	.0010	.8656	.2919	-.0619
10	.5404	-.0323	.1575	.6569	.2881

注：抽出後 Varimax 回転したもの

全て収束が一意でなかった。m=2の方は $\frac{P-1}{2} \geq m$ でも収束が一意でない場合がありうることを示すものである。

Reference

1. Cattell, R.B. (1965): Factor Analysis - An Introduction to Essentials, I. The Purpose and Underlying Models. Biometrics.
2. Cliff & Pennell (1967): The Influence of Communality, Factor Strength and Loading Size on the Sampling Characteristics of Factor Loadings. Psychometrika, vol.32 No.3.
3. Guttman, L. (1958): What Lies Ahead for Factor Analysis? Educational and Psychological Measurement. vol.18. No.3.
4. Jöreskog, K.G. (1963): Statistical Estimation in Factor Analysis. Stockholm. Almquist and Wiksell.
5. Rao, C.R. (1964): The Use and Interpretation of Principal Component Analysis in Applied Reserch. Technical Report No9.
6. Thurstone, L.L. (1947): Multiple Factor Analysis. Chicago, Univ. of Chicago Press.
7. Wold, H. (1953): Some Artificial Experiments in Factor Analysis. Uppsala Symposium on Psychological Factor Analysis, Nordisk Psychologi's Monograph Series No3. Almquist and Wiksell.